ARL-TN-0684 ● JULY 2015

US Army Research Laboratory

Monte Carlo Simulations: Number of Iterations and Accuracy

by William Oberle

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TN-0684 ● JULY 2015

US Army Research Laboratory

Monte Carlo Simulations: Number of Iterations and Accuracy

by William Oberle
Weapons and Materials Research Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
July 2015	Technical Note	1 February–30 April 2015

4. TITLE AND SUBTITLE
Monte Carlo Simulations: Number of Iterations and Accuracy

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
William Oberle

5d. PROJECT NUMBER
AH80

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL-WML-A
Aberdeen Proving Ground, MD 5066

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TN-0684

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report focuses on 2 related topics, the number of Monte Carlo iterations and the accuracy or error in the estimation of the mean of the probability distribution for the quantity of interest being analyzed. For most Monte Carlo simulations, it is the estimation of this mean that is desired. These 2 topics are related through the central limit theorem, and given one, the other can be determined when combined with the sample information for large sample sizes. Issues associated with small sample size have been mentioned and discussed in more detail for quantities of interest with a normal distribution.

15. SUBJECT TERMS
Monte Carlo, confidence interval, central limit theorem, number of iterations, Wilson score method, Wald method, normal probability plot

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			William Oberle
Unclassified	Unclassified	Unclassified	UU	34	19b. TELEPHONE NUMBER (Include area code) 410-278-4362

Contents

List of Figures

List of Tables

Acknowledgments

The author would like to thank Ms Mary Arthur, Mr Andrew Thompson, and Dr Benjamin Breech for their technical review of this report, and Ms Jessica Schultheis for her editorial review.

INTENTIONALLY LEFT BLANK.

1. Introduction

Monte Carlo (MC) methods[1] are often used when a closed-form solution for a property being studied cannot be developed. Mathematically, we are attempting to evaluate $\int_{\Omega} p(\cdot)F(\cdot)d\Omega$, where $p(\cdot)$ is the combined probability function of all the input parameters, $F(\cdot)$ is the function describing the property being studied, and Ω is the domain of interest (e.g., the 2-dimensional plane of the battlefield in a weapons analysis). If we cannot evaluate this definite integral, numerical methods must be used. MC methods are a set of numerical methods that are especially useful when dealing with probability distributions.

The defining characteristic of MC methods is the random generation of input parameter values from probability distributions. For example, consider determining the value of the mathematical constant π. A straightforward geometric approach is shown in Fig. 1, where a circle of diameter D is inscribed in a square with side length D. Computing the ratio of the area of the 2 figures,

$$\frac{Area\ \ Circle}{Area\ \ Square} = \frac{\pi D^2/4}{D^2} = \pi/4, \tag{1}$$

eliminates D and gives a value for π after multiplying by 4. A numeric or closed-form solution for the ratio of the areas is not possible, so the use of an MC method to approximate π is appropriate.

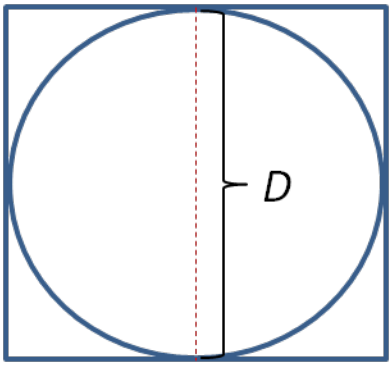

Fig. 1 Geometric domain for approximating π

An MC method/algorithm to approximate π follows.

1) Select 2 random numbers, x_1 and x_2, from the interval $[0,D]$ (random generation of input parameter values).

2) Determine if the point defined by the ordered pair (x_1,x_2) lies within or on the circle. Keep track of the total number of points within or on the circle and the total number of points tested.

3) Approximate the ratio of the areas by the number of points within or on the circle divided by the total number of points tested.

4) Approximate π by multiplying the value in step 3 by 4.

5) Repeat steps 1–4 until π is approximated to the desired accuracy.[2] The number of times steps 1–4 is repeated is the number of iterations.

Table 1 presents results for the approximation of π using the algorithm describe above for different numbers of iterations along with the percent difference between the approximation and π.

Table 1 Approximation of π using the MC method $\pi \approx 3.14159265$

Number of Iterations	Approximation to π	Percent Difference
100	3.4	8.2254
1,000	3.104	–1.1966
10,000	3.1608	0.6114
100,000	3.14756	0.1899
1,000,000	3.14260	0.0321
10,000,000	3.1412788	–0.0100
100,000,000	3.14155688	–0.0011

As expected, the approximation of π becomes better as the number of iterations increases, and we would argue that the result is accurate to 3 decimal places. We expect the accuracy to increase as the number of iterations increases. Fortunately, as will be shown later, if the MC simulation is properly formulated, this is statistically true with the potential error in the approximation being proportional to $1/\sqrt{n}$ with n being the number of iterations. This guarantees that for any MC simulation there is a calculable number of iterations to be performed that will provide an approximation to any desired accuracy. However, it does not tell us how many iterations should be performed. Or, more importantly, how accurate the resulting estimation is. These are important issues that the analyst should address before starting the analysis. Yet, Robey and Barcikowski[3] cite a report by Hauck and Anderson[4] stating that "in a survey of simulation studies, [they] found that only 9 percent of the surveyed reports included a justification for the number of iterations utilized".

The object of this report is to address statistical approaches to both of these questions. The remainder of this report is organized as follows. In Section 2, the central limit theorem[5] (CLT) is discussed. The third section provides several approaches for estimating the number of MC iterations required to achieve a desired accuracy. Section 4 discusses estimating MC simulation result accuracy for both a large and small number of iterations. The use of the percentage error of the mean to estimate the number of MC iterations and accuracy is presented in Section 5. The final section provides conclusions.

2. Central Limit Theorem

The statistical foundation for the work presented in this report is the CLT. Before stating the theorem, some comments on notation are needed. Let Y represent a population with a distribution that has a mean of μ and a variance σ^2. No assumption is made about Y's distribution—it may or may not be a normal distribution. Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample with replacement of size n drawn from the Y population. If we take the average of this sample, $\bar{y} = (\Sigma Y_i)/n$, we produce a single point estimate for the mean of the Y distribution. There is another population with its own distribution known as the distribution of the sample mean. This population consists of all estimations of the Y distribution's mean possible by averaging random samples of size n drawn with replacement from Y. Denote this distribution by $\bar{Y}$. Thus, $\bar{y}$ is an element of $\bar{Y}$. Since $\bar{Y}$ is a population, it has a mean and variance denoted, respectively, by $\mu_{\bar{Y}}$ and $\sigma^2_{\bar{Y}}$. The CLT gives the relationship between the mean and variance of the Y and $\bar{Y}$ distributions. Devore[5] defines the CLT as follows:

> CLT: Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a distribution Y with mean μ and variance σ^2. Then if n is sufficiently large, $\bar{Y}$ has approximately a normal distribution with $\mu_{\bar{Y}} = \mu$ and $\sigma^2_{\bar{Y}} = \sigma^2/n$. The larger the value of n, the better the approximation.

To illustrate the CLT, we return to the MC algorithm for approximating π described in the previous section. The algorithm relied on taking the average of a number of samples drawn from the probability distribution defined by performing a single iteration of the algorithm. For a single iteration, if the selected point is within or on the inscribed circle, the ratio of the areas would equal one, thereby producing a value of 4 since the ratio of the areas is multiplied by 4 in the algorithm. If the point is not within or on the circle, the result is zero since the ratio of the areas is zero and multiplying by 4 is still zero. The probability density function (pdf) can be determined using geometric probabilities.[6] So the probability that a random point

is within or on the circle is π/4, and outside the circle the probability is 1 – π/4. Formally, the pdf is

$$p(y) = \begin{cases} \dfrac{\pi}{4} & \text{if y = 4, i.e., a point within or on the circle} \\ 1 - \dfrac{\pi}{4} & \text{if y = 0, i.e., a point outside the circle .} \end{cases} \quad (2)$$

This pdf is the transformation of a binomial distribution. In terms of the CLT, this is the Y distribution. The terms "sample size" and "number of iterations" are used somewhat interchangeably. For example, if we perform *n* iterations, we have a sample of size *n*.

Applying the standard formula for the mean (expected value) of a distribution, $\mu = E(Y) = \sum_{y \in Y} y \cdot pdf(y)$, yields

$$\mu = 4 * \pi/4 + 0 * (1 - \pi/4) = \pi. \quad (3)$$

For future reference, notice that estimating π is equivalent to estimating or determining a characteristic of the Y distribution—specifically, we seek a numerical estimate of the mean of the Y distribution.

Using the alternate formula for variance, $\sigma^2 = E(Y^2) - [E(Y)]^2$ produces

$$\sigma^2 = [16 * \pi/4 + 0 * (1 - \pi/4)] - \pi^2 = 4\pi - \pi^2. \quad (4)$$

In the MC simulation for this example, a sample with a single iteration would be a value of 0 or 4.

If we were to perform 2 iterations and compute the average of the 2 samples, we would produce the distribution shown in Table 2. This is the CLT with *n* = 2 (i.e., a sample size of 2).

Table 2 Distribution information for the average of 2 samples

Sample 1	Sample 2	Average of Sample 1 and Sample 2	Probability of Average
0	0	0	$(1 - \pi/4)^2$
0	4	2	$(1 - \pi/4) * \pi/4$
4	0	2	$\pi/4 * (1 - \pi/4)$
4	4	4	$(\pi/4)^2$

There are now 3 possible outcomes: 0, 2, and 4. If we were to use 2 iterations in a MC simulation and take the average, we would have a single value consisting of one of these values. Computing the mean and variance for the average of 2 samples yields the following results:

$$\mu = 0 * (1 - \pi/4)^2 + 2 * [2 * \pi/4 * (1 - \pi/4)] + 4 * (\pi/4)^2$$

$$= \pi * (1 - \pi/4) + \pi^2/4 = \pi - \pi^2/4 + \pi^2/4 = \pi \qquad (5)$$

and

$$\sigma^2 = 0^2 * (1 - \pi/4)^2 + 2^2 * [2 * \pi/4 * (1 - \pi/4)] + 4^2 * (\pi/4)^2 - \pi^2$$

$$= 2\pi - \pi^2/2 + \pi^2 - \pi^2 = 2\pi - \pi^2/2 = (4\pi - \pi^2)/2. \qquad (6)$$

Eqs. 5 and 6 are the values predicted by the CLT for $\mu_{\bar{Y}}$ and $\sigma^2_{\bar{Y}}$ with a sample size of 2. This illustrates the predicted CLT values for the mean and variance of the distribution of the sample mean. To illustrate that the distribution of the sample mean tends toward a normal distribution, we will consider averages of samples of sizes 10 and 20. Figure 2 shows the pdf for the base distribution, Eq. 3, and sample sizes of 2, 10, and 20.

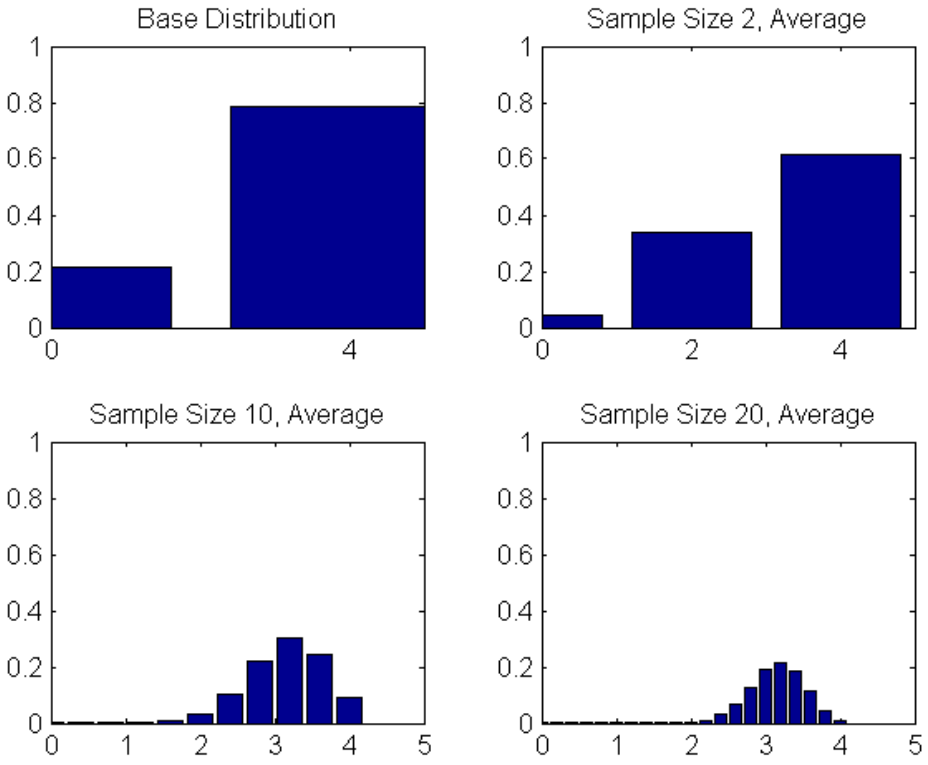

Fig. 2 Pdf for the base distribution and for sample sizes of 2, 10, and 20

The trend toward a normal distribution with mean π is evident in the 4 pdfs. In all cases the computed mean for the pdf is π and the variances, in order, are 2.696766213, 1.3483831, 0.2696766213, and 0.1348383106. The variance values are in agreement with the CLT (i.e., $(4\pi - \pi^2)/n$ for n = 1, 2, 10, and 20). As the sample size approaches infinity, the $\bar{Y}$ distribution's variance will approach zero. In effect, the $\bar{Y}$ pdf is approaching the Dirac delta function centered at the mean of the original distribution.

In general, how does the CLT apply to MC methods and simulations? In an MC simulation, values for the input parameters are randomly generated based upon knowledge of the pdf associated with each input parameter.[7] These values serve as the inputs to a deterministic model/process to produce the answer for that particular set of input parameters. Since there is most likely variability in at least some of the input parameters, there is no single answer to the problem being studied; rather, there is a probability distribution characterizing the totality of the possible answers. In an MC simulation, we are attempting to determine the characteristics that define this distribution; this is the Y distribution mentioned in the CLT. The CLT is the theoretical tool that tells us how to determine the mean of the Y distribution and at the same time bound the confidence interval (CI) for the mean. The CLT also provides justification for an a priori (before running the MC simulation) estimate of the number of iterations to be performed by the MC simulation to provide a desired level of accuracy for the results.

Suppose that we perform an MC simulation with *n* iterations. This means that we have drawn a sample of size *n*, $Y_1, Y_2, \ldots, Y_n$, from the Y distribution or population. There is a related population made up of the average values for every possible sample of size *n*, the $\bar{Y}$ distribution in the CLT. Thus, the output of our MC simulation can be thought of as a sample of size *n* from the Y distribution or a sample of size 1 from the $\bar{Y}$ distribution for sample size *n*. Since the goal of the MC simulation is to define characteristics of the Y distribution, why not simply use the sample of size *n* to determine the desired characteristics of Y? The simple answer is that we can estimate the mean of the Y distribution with the sample average, but we have no way to estimate the accuracy or error bound on the mean. Fortunately, the CLT guarantees that $\mu_{\bar{Y}} = \mu$, $\sigma^2_{\bar{Y}} = \sigma^2/n$, and the $\bar{Y}$ distribution asymptotically approaches the normal distribution. Thus, the $100*(1 - \alpha)$ % confidence interval is given by[8]

$$Y_{average} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}, \tag{7}$$

where $Y_{average}$ = average $(Y_1, Y_2, \ldots, Y_n)$ and *s* = standard deviation $(Y_1, Y_2, \ldots, Y_n)$ are used as the unbiased estimators for the Y distribution μ and σ. $z_{\alpha/2}$ is the

standard normal distribution z-score such that α/2 is the area under the standard normal curve to the right of $z_{\alpha/2}$. To compute the number of iterations to bound the half-width of the CI, we would need to know *s*. But *s* is unknown until a number of iterations have been performed; thus, the number of iterations to achieve a specific bound on the half-width of the CI cannot be estimated a priori. In those cases where the variance of the Y distribution is known, σ can be substituted for *s* and an a priori calculation for *n* is possible. As will be seen in the next section for the special case where Y has a binomial distribution, an upper bound on the number of MC iterations required for a given accuracy can be determined.

Equation 7 does provide a condition that allows the MC simulation to be terminated when the desired half-width of the CI, Δ (i.e., the desired accuracy for the mean of the Y distribution) has been achieved. Specifically,

$$z_{\alpha/2}\frac{s}{\sqrt{n}} < \Delta. \tag{8}$$

We will return to Eq. 8 in the next 2 sections.

3. A Priori Estimate of Number of MC Iterations

Over the past several decades, analysts with the Advanced Lethality and Protection Analysis Branch (ALPAB) of the Weapons and Materials Research Directorate, US Army Research Laboratory, have grappled with questions regarding the number of iterations and the accuracy of results when performing their MC simulations. Several different approaches relying on the CLT and CI for a population proportion have been used. Since the majority of the ALPAB analyses involved a personnel incapacitation measure of performance, the idea of using a proportion (incapacitated or not incapacitated) was felt to be justified.

The derivation of the CI for a population proportion, *p*, based upon a random sample of size *n*, can be found in almost every text on introductory statistics, such as Devore,[9] and is given in Eq. 9. Here $\hat{p}$ is the natural estimator of *p* and is given by the fraction of successes, *X/n*, from a random sample of size *n* drawn from the population; *X* is the number of successes; $\hat{q} = 1 - \hat{p}$; α is the level of confidence, generally 95% or 99%; and $z_{\alpha/2}$ is the z-score associated with α/2.

$$P\left(-z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{\hat{p}\hat{q}/n}} < z_{\alpha/2}\right) \approx 1-\alpha. \tag{9}$$

The most common method used to determine CIs and the number of iterations for population proportion is the method by Wald[10] (WM) that assumes the probability in Eq. 9 is equal to 1 – α and yields the formula

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{\Delta^2} \tag{10}$$

for the number of iterations. Here, Δ is half the length of the confidence interval. Since we are talking about a population proportion, $0 < p < 1$, $\Delta = 0.005$ would represent a CI of width 1% (i.e., if $\hat{p} = 0.58$ or 58%, a $\Delta = 0.005$ would give a CI between 57.5% and 58.5%, a width of 1%). Analysts with ALPAB used Eq. 10 to obtain an estimate for the number of iterations to use in the MC simulation by letting $\hat{p} = 0.5$, the value that would maximize the right-hand side of Eq. 10. For example, with $\hat{p} = 0.5$, Eq. 10 predicts that 960,400 iterations are required for a 95% CI ($z_{\alpha/2} = 1.96$) with $\Delta = 0.001$.

Returning to our estimation of π, we were estimating a proportion in calculating the ratio of the areas of the circle and square. To estimate π, we multiplied the calculated proportion by 4. Thus, if the half-width of the CI for the proportion is Δ, the half-width for the estimation of π will be 4*Δ. From Table 1, π was estimated to be 3.14260 for 1,000,000 iterations, and we would estimate the 95% CI to be approximately 3.14260 ± 0.004 or (3.13860, 3.14660). "Approximately" is used here because we would actually expect the CI to be slightly smaller since 1,000,000 iterations, not 960,400, were used in the approximation of π. The difference between π and 3.14260 is approximately –0.00100, well within the ±0.004.

Unfortunately, such good results are not always observed for our MC simulations, especially when *p* is close to 0 or 1. As stated by Dunnigan,[11]

> Careful study however reveals that it [Wald method] is flawed and inaccurate for a large range of n and p, to such a degree that it is ill-advised as a general method.[12,13] Because of this many statisticians have reverted to the exact Clopper-Pearson method, which is based on the exact binomial distribution, and not a large sample normal approximation (as is the Wald method). Studies have shown however that this confidence interval is very conservative, having coverage levels as high as 99% for a 95% CI, and requiring significantly larger sample sizes for the same level of precision.[12–14] An alternate method, called the Wilson Score method is often suggested as a compromise. It has been shown to be accurate for most parameter values and does not suffer from being over-conservative, having coverage levels closer to the nominal level of 95% for a 95% CI.

Agresti[15] further quantifies Dunnigan's statement concerning the WM, stating that the method is inaccurate for small values of *n* and when *p* is close to 0 or 1 (extreme values).

As a result of the critique of Wald's method by Dunnigan and Agresti, as well as our own observations, we opted to use the Wilson score method[16] (WSM). The WSM confidence interval is shown in Eq. 11.

$$\left(\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \left(z_{\alpha/2}^2\right)/n}, \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \left(z_{\alpha/2}^2\right)/n} \right). \qquad (11)$$

We developed a procedure using Eq. 11 to determine when to terminate a simulation based upon the CI half-length being less than Δ. The procedure consisted of determining if the inequality $\frac{z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \left(z_{\alpha/2}^2\right)/n} < \Delta$ is true at the completion of each MC iteration. If the inequality is true, the MC simulation is terminated, otherwise the MC simulation would continue. However, this procedure was not very useful. The number of iterations that needed to be performed was sensitive to the value of Δ, often resulting in a large number of iterations being performed. This was especially true during the initial MC simulations performed during an analysis when the value of $\hat{p}$ would not be well know. Use of the procedure was quickly discontinued.

Although more complicated than the WM, the WSM can also be solved for the upper bound on the number of required iterations since $\hat{p} = 0.5$ also maximizes the expression for the half-width in Eq. 11. Using the WSM with the same values used earlier in the WM to estimate the number of iterations needed to achieve a half interval of 0.001 length gives 960,396 iterations versus the 960,400 iterations predicted previously using the WM. The difference of 4 iterations out of almost 1,000,000 iterations is virtually meaningless and calls into question if the WSM is useful in estimating a priori the number of iterations because of its added complexity compared to the WM. We recommend that the WM be used for a priori estimates of the number of MC iterations since the maximizing value for the half-width of the WM CI in Eq. 10 is $\hat{p} = 0.5$, and this value is not close to 0 or 1 where the WM is inaccurate.[15]

Although the WM and the WSM have generally proven useful in estimating the number of MC iterations and addressing the accuracy of the MC simulation results, these methods do have drawbacks. First, they both assume that the random variable

being studied is a binomial random variable. Fortunately, this is the case for many analyses if they are thought of "in the right way". Consider our estimation of π using the algorithm above. The objective of the algorithm, estimating π, was not a binomial experiment. The binomial experiment was the counting of the number of times that the randomly chosen point was within the circle. This binomial random variable directly leads to the estimation of π when we multiplied the binomial random variable by 4.

This transformation of the binomial random variable (i.e., multiplying by 4) leads to the second drawback: the measured accuracy of the variable being estimated by the MC simulation. In our example, we set the desired width of the half interval to be 0.001. But as illustrated above, this was not the precision of the estimation for π. It was the half-width of the CI for the binomial random variable, the probability of a randomly chosen point being within the circle. The precision of the estimation for π had to undergo the same transformation as the transformation applied to the binomial random variable to estimate π (i.e., multiply by 4). Thus, the half-width for the 95% CI for the accuracy of the estimation for π was 0.004.

It is not difficult to keep track of the transformations between the random variable and adjust the CI half-widths so that the number of iterations or desired CI are correctly computed; however, it is possible to eliminate the dependency on requiring a binomial random variable and the added complexity. The cost of removing the requirement of a binomial random variable is that we cannot make a priori estimates of the number of iterations required to achieve a desired level of accuracy without additional information that will probably not be known. What the binomial random variable provided was a value for $\hat{p}$, 0.5, that provided an upper bound for the variance needed to calculate the number of iterations.

Equation 8 can be solved for *n* to obtain a result similar to Eq. 10,

$$n = \frac{z_{\alpha/2}^2 s^2}{\Delta^2}. \tag{12}$$

If the variance, σ^2, of the Y distribution is known, it can be used in place of s^2 in Eq. 12, and the value of *n* can be determined. If the variance is not known, the standard deviation of a small sample could be used to estimate *n*. As more iterations are performed, refined estimates for *n* can be obtained until the estimates for *n* converge.

Table 3 summarizes the results of using this iterative approach to estimate the number of iterations required to estimate π to within an accuracy of 0.004 and 0.001 based upon the information in Table 1. The actual variance for the base or Y distribution is $4\pi - \pi^2 = 2.696766213$ (see Eq. 4). Using this value for s^2 in Eq. 12

with $z_{\alpha/2}$ = 1.96 (95% confidence level) yields n values of 647,494 and 10,359,897 for Δ = 0.004 and 0.001, respectively. The estimated number of iterations in Table 3 appears to be converging to these values as the sample size increases. Also, for a sample size of 1,000 (1,000 iterations) the estimated number of iterations for both levels of accuracy, Table 3, is only about 3% larger than the number of iterations calculated using the true value for σ in Eq. 12 (i.e., 668,431 vs. 647,494 for Δ = 0.004 and 10,694,891 vs. 10,359,897 for Δ = 0.001).

Table 3 Estimation of the number of MC iterations required to achieve a half-width of 0.004 and 0.001 for the 95% level of confidence (variance of the Y population = 2.696766213)

Sample Size	Sample Variance (Estimation of Y Population Variance)	Estimated Number of Iterations Half-Width = 0.004	Estimated Number of Iterations Half-Width = 0.001
100	2.060606	494,752	7,916,024
1,000	2.783968	668,431	10,694,891
10,000	2.652809	636,939	10,191,031
100,000	2.683133	644,220	10,307,524
1,000,000	2.694466	646,941	10,351,061

The WM or the WSM a priori estimate for the number of iterations required to achieve 0.004 accuracy in estimating π was approximately 960,400. As illustrated above, the iterative method using Eq. 12 to achieve the same accuracy indicates that approximately 650,000 iterations are required. Thus, the iterative method predicts about two-thirds of the number of iterations predicted by population proportion, WM or WSM, methods. As mentioned above, the WM and WSM provide an upper bound for the number of iterations since $\hat{p} = 0.5$ maximized the iteration estimate given by Eqs. 10 and 11, so the fewer number of required iterations predicted by the iterative method is not unexpected.

4. MC Result Accuracy

In Section 2 we stated that we are attempting to determine the characteristics that define the Y distribution mentioned in the CLT. If we are trying to determine the mean, μ, of the Y distribution, the CLT provides 2 approaches for determining the accuracy of our result. In the previous section, we discussed methods for selecting the number of MC iterations to guarantee a user-specified level of accuracy for μ. Unfortunately, the required number of iterations could be prohibitive in actual practice, most often due to time constraints. This brings us to the second approach that is the classical statistical methodology for CIs. Assuming n is sufficiently large, generally taken as $n \geq 30$, the CLT states that $\bar{Y}$ has an approximately normal distribution no matter the Y population distribution and[17]

$$P\left(-z_{\alpha/2} < \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \approx 1-\alpha. \quad (13)$$

Note that Eq. 9 is a special case of Eq. 13, applicable to the binomial distribution. Solving the inequality for μ produces a CI for μ with a confidence level approximately 100*(1 – α) % and is shown in Eq. 14,

$$\overline{Y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{Y} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \quad (14)$$

Estimating $\bar{Y}$ with the sample average, $Y_{average}$, and σ with the sample standard deviation, *s*, gives the final equation for the CI,

$$Y_{average} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < Y_{average} + z_{\alpha/2}\frac{s}{\sqrt{n}}. \quad (15)$$

Using the information in Tables 1 and 3 with Eq. 15, we find the 95% confidence levels for the sample average for sample size *n* (Table 4).

Table 4 Confidence intervals at the 95% level of confidence for the estimation of π by the sample average for different sample sizes together with the length of the half-width of the confidence interval

Sample Size	Sample Variance (Estimation of Y Population Variance)	Sample Average (Estimation of Y Population Average)	Lower Bound 95% Confidence Level for Mean	Upper Bound 95% Confidence Level for Mean	Length of Half-Width
100	2.060606	3.4	3.118645704	3.681354296	0.281354296
1,000	2.783968	3.104	3.000583892	3.207416108	0.103416108
10,000	2.652809	3.1608	3.128876606	3.192723394	0.31923394
100,000	2.683133	3.14756	3.137407402	3.157712598	0.10152598
1,000,000	2.694466	3.1426	3.139382694	3.145817306	0.0032173.6

At this point we have 3 different ways to estimate CIs for the mean using a sample size of *n*. Besides Eq. 15, there are the 2 approaches under the assumption that the final distribution is the transform of a binomial distribution through the use of Eqs. 9 and 11. A comparison of the results for these 3 methods is given in Table 5 for the sample size of 1,000,000.

Table 5 Confidence intervals at the 95% level of confidence for the estimation of π using the three methods discussed in the paper, sample size = 1,000,000

Method	Sample Average	Lower Bound 95% Confidence Level for Mean	Upper Bound 95% Confidence Level for Mean	Length of Half-Width
Wald: Eq. 9	3.14260	3.13938269	3.14581731	0.00321731
WSM: Eq. 11	3.14259561	3.13937831	3.14581291	0.00321730
CLT: Eq. 15	3.14260	3.13938269	3.14581731	0.00321731

All 3 estimates are essentially the same with the WSM providing a slightly better estimate for π because of the correction of the average in Eq. 11. When the transformation (i.e., multiply by 4) is used, the binomial distribution standard deviation to use in Eq. 15 would be

$$4\sqrt{\hat{p}\left(1-\hat{p}\right)} = 4\sqrt{\frac{3.1426}{4}\left(1-\frac{3.1426}{4}\right)} = 4\sqrt{0.78565*0.21435} = 4\sqrt{.168404} \quad (16)$$
$$= 4*0.4103706 = 1.641483.$$

From Table 4, the computed sample standard deviation for a sample size of 1,000,000 is $\sqrt{2.694466} = 1.641483$, which explains the same results between the 3 methods.

The CI calculations discussed in this section have assumed that the population distribution was a binomial distribution or that the sample size was sufficiently large so that the CLT could be used regardless of the nature of the population distribution. As mentioned earlier, a large sample size is considered to be $n \geq 30$. Unfortunately, many of today's high-fidelity physics-based performance models are time intensive, and performing 30+ MC iterations is prohibitive, resulting in a small sample size. For small sample sizes, a formula for the CI of the population mean can still be developed. This small sample size CI is tailored to the assumed population distribution (normal, gamma, etc.) and may be different for each population distribution assumption.

For example, if the population distribution is assumed to be a normal distribution, the resulting formula for the CI of the population mean is based upon the *t*-distribution if the population standard deviation is estimated by *s*. This CI for μ with a confidence level approximately 100*(1 – α) % is shown in Eq. 17.[18]

$$Y_{average} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < Y_{average} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}. \quad (17)$$

In Eq. 17, $t_{\alpha/2,n-1}$ is the *t*-critical value at the 100*(1 – α) % level of confidence with $n - 1$ degrees of freedom (DoF). As $\text{DoF} \to \infty$, the distribution approaches the normal distribution.[17] For $n = 30$, the *t*-critical value, 2.042, is approximately 4% higher than the normal value, 1.96, for the 95% confidence level.

For small sample sizes when there is no a priori knowledge of the population distribution, the CLT cannot be used to assume that the population distribution is approximately normal and the CI estimates discussed above may not be valid. In these circumstances, testing is necessary to determine if the sample could be from a normal distribution. Several approaches are available to assess if the sample could be from a normal distribution so that Eq. 17 could be used to estimate a CI for the population mean when dealing with small sample sizes. There are a number of statistical tests for normality available in most statistical software packages. Wikipedia summarizes several of these normality tests[19] as follows.

> Tests of univariate normality include D'Agostino's K-squared test, the Jarque–Bera test, the Anderson–Darling test, the Cramér–von Mises criterion, the Lilliefors test for normality (itself an adaptation of the Kolmogorov–Smirnov test), the Shapiro–Wilk test, the Pearson's chi-squared test, and the Shapiro–Francia test. A 2011 paper from The Journal of Statistical Modeling and Analytics[20] concludes that Shapiro-Wilk has the best power for a given significance, followed closely by Anderson-Darling when comparing the Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests.

Having used the Anderson-Darling and the Shapiro-Wilk tests, I recommend that care be taken when interpreting the results. Before any statistical test is used, a normal probability plot[21] (NPP) should be constructed and analyzed. If the NPP indicates normality, one of the statistical tests for normality can be performed to quantify the confidence level of a normality assumption.

The basic idea of an NPP is to plot the sample data in such a way that if the points fall on a straight line, the sample data were most likely randomly chosen from a population with normal distribution.[22] Deviation from a straight line indicates that the population distribution is not normal, and how the points deviate from the straight line provides some insight into the true population distribution. For a discussion on interpreting NPPs, see, for example, Normal Probability Plot on the National Institute of Standards and Technology website.[23]

Devore[19] presents one of a variety of approaches for constructing an NPP that does not require the sample data to be transformed. His method relies on the definition of sample percentiles:

Order the n sample observations from smallest to largest. Then the i^{th} smallest observation in the list is take to be the $[100(i - 0.5)/n]$th sample percentile.

If the population from which the sample was drawn has a normal distribution, the sample percentile for a data point should match the normal z-percentile. The normal z-percentile is defined to be the z-value of the normal distribution at which the probability of the standard normal random variable will be less than or equal to a given probability. For example, the z-percentile for 0.025 is –1.96 (i.e., the value from the standard normal table for which the area to the left is 0.025. Plotting the $(i - 0.5)/n$ z-percentile versus the i^{th} smallest sample data point produces the NPP. MATLAB code for creating this NPP is provided in the Appendix.

Figure 3 is the NPP based upon a sample size of 30 drawn from a normal distribution, with $\mu = 10$ and $\sigma = 3$. The sample mean is 9.823 and the sample standard deviation is 3.3513. The red line in the figure is based upon Eq. 18 and is valid for the normal distribution with $\mu = Y_{average}$ and $\sigma =$ sample standard deviation.

$$Observation = s * z_{percentile} + Y_{average}. \tag{18}$$

The plotted points appear to form roughly a straight line indicating the sample was most likely drawn from a normal population. This conclusion was supported by using the Anderson-Darling test on the sample data. Results are not shown.

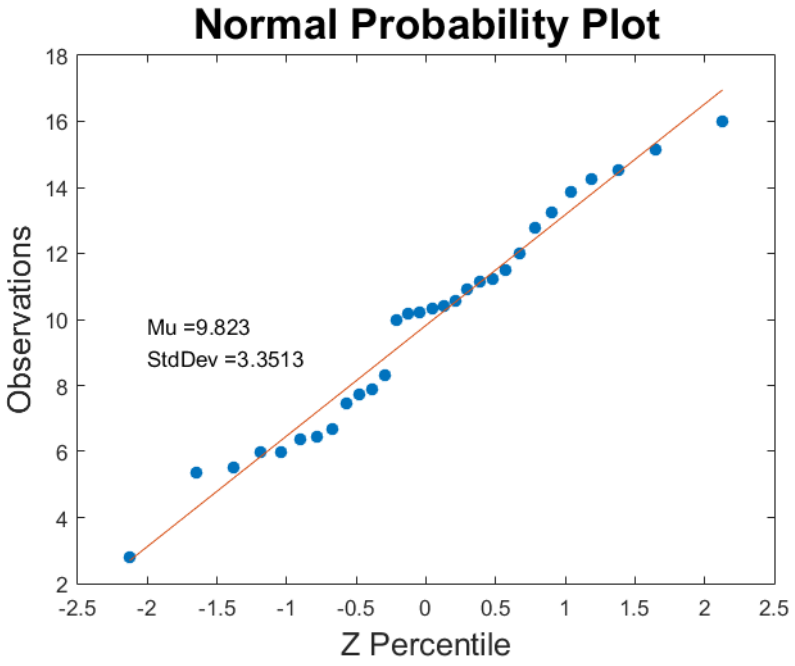

Fig. 3 NPP for a sample of size 30 drawn from a normal distribution with $\mu = 10$ and $\sigma = 3$

5. Using Percentage Error of the Mean to Estimate Number of MC Iterations

As seen in the previous sections, the a priori estimation of the number of iterations to perform in an MC simulation to achieve the desired accuracy for the result was rather large. A large number of iterations was also required to achieve narrow CIs for the mean. Statistically, there is no way around this predicament unless we reduce the desired accuracy. This is seen from Eq. 12 since the sample standard deviation, s, and the z-score, $z_{\alpha/2}$, are fixed, leaving only an increase in the CI half-width, Δ, as the only way to significantly reduce n. Driels and Shin[25] recommend using the percentage error of the mean instead of the half-width of the CI. Essentially, this changes how we look at accuracy from an absolute (e.g., accurate to a specific number of decimal places) to what fraction of the true answer is our result. For example, for 1,000 iterations our estimation of π, Table 1, was accurate to within ≈ 0.0376 or 1.2%. Being accurate within less than 2 decimal places does not sound as good as being within 1.2% of the answer. Another advantage of this approach is that percentage error is a normalized value, and we do not have to choose accuracy based upon the expected true value. The percentage error of the mean can be found starting with Eq. 15. Subtracting $Y_{average}$ gives

$$-z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu - Y_{average} < z_{\alpha/2}\frac{s}{\sqrt{n}}. \tag{19}$$

Dividing by $Y_{average}$ produces

$$-z_{\alpha/2}\frac{s}{Y_{average}\sqrt{n}} < \frac{\mu - Y_{average}}{Y_{average}} < z_{\alpha/2}\frac{s}{Y_{average}\sqrt{n}}. \tag{20}$$

The middle term is the fraction of error for the mean, and multiplying through by 100 will convert to percentage,

$$-z_{\alpha/2}\frac{100s}{Y_{average}\sqrt{n}} < \frac{100\left(\mu - Y_{average}\right)}{Y_{average}} < z_{\alpha/2}\frac{100s}{Y_{average}\sqrt{n}}. \tag{21}$$

Thus, the maximum percentage error of the mean, denoted by ε, is the right-most expression[26]

$$\varepsilon = z_{\alpha/2}\frac{100s}{Y_{average}\sqrt{n}}. \tag{22}$$

Solving for n produces an estimate for the number of iterations required to achieve a percentage error of the mean equal to ε.[27]

$$n = \left[z_{\alpha/2} \frac{100s}{\varepsilon Y_{average}} \right]^2 . \tag{23}$$

As with the application of Eq. 15, the sample average, $Y_{average}$, and sample standard deviation, s, must be estimated using some initial sample size. Since the CLT is the basis for this estimate, the initial sample size should be greater than or equal to 30. As stated earlier, as more iterations are performed, refined estimates for n can be obtained until the estimates for n converge.

To illustrate the use of this approach, suppose we wish to use our MC simulation for π to approximate π to within 2%. When the values from Table 4 are used with a 95% level of confidence for 100 iterations, Eq. 23 predicts that 1,712 iterations would be required. After 1,000 iterations we could recalculate the estimate for n. Again, using the values in Table 4, we find that the estimated number of iterations would be 2,775. When the exact values are used for $Y_{average}$ and s (Eqs. 3 and 4), π and $4\pi - \pi^2$ yield the exact number of iterations, 2,550. So if the MC simulation is run 2,550 times, we are 95% confident that the calculated value for π is within 2% of the true value of π. Since 2,775 exceeds 2,550, we would be closer than 2% if we did no update of the required number of iterations after 1,000 iterations.

6. Conclusions

From a statistical standpoint, the goal of MC simulation methodology is to determine the characteristics of the probability distribution associated with a random variable describing a real-world quantity of interest (QI) to the researcher. We can think of the MC method as consisting of 4 distinct components.

1) A deterministic model that calculates a value for the QI given a set of values for the input parameters to the model.

2) A process for selecting a set of values for the input parameters given sufficient information for the parameters. If there is no uncertainty in any of the input parameters, the unique value for each input parameter must be known. In this case, an MC simulation is not required, or we can think of this case as a trivial MC simulation with a single iteration and the probability distribution consists of a single point with probability 1.

3) A loop that repeats 1 and 2. This component constructs a sample size of n for the population of the QI random variable.

4) A statistical analysis of the sample from 3 to estimate characteristics of the probability distribution for the population of the QI random variable. Generally, we are interested in the type of distribution (e.g., normal, gamma) and its defining characteristics (e.g., mean and variance for a normal distribution).

In terms of the number of iterations, providing an a priori estimate of the number of iterations for the loop to guarantee a specified level of accuracy for the mean would be the goal. However, this is not possible unless additional information about the probability distribution of the random variable associated with the QI is known (specifically, the type of distribution and any parameters necessary to determine the width of CIs for a given level of confidence in that distribution). If the distribution is known to be normal, the variance is needed. The closest one can come to an a priori estimate for the number of iterations if the probability distribution is binomial or a linear transformation of a binomial distribution. For these distributions, $\hat{p} = 0.5$ will maximize the variance, and an upper bound on the number of iterations can be estimated using the WM (Eq. 10) or the WSM (Eq. 11). Since both equations will give essentially the same result for large n, using the WM is recommended because of its simplicity. The WSM should be used when n is small or $\hat{p}$ is close to 0 or 1 because of its better accuracy in these cases. In all cases, the WSM will provide a slightly more accurate estimation of the mean. An iterative approach to determining the number of iterations based upon the CLT can also be used. This approach does not provide an a priori estimate of the number of iterations but does provide a method for terminating the MC simulation while ensuring a desired level of accuracy for the QI.

The accuracy of the estimation of the mean can be determined using Eq. 15, which is based upon the CLT and the properties of the normal distribution. This equation is valid no matter the type distribution of the random variable for the QI as long as the sample size is sufficiently large. Most statistical texts use a sample size of 30 as the boundary between small and large samples. For small samples, the *t*-distribution and Eq. 17 can be used to determine a CI for the population under the assumption that the population is normal. When the analyst is faced with a small sample size and uncertainty about the population distribution, NPP and statistical tests, such as the Anderson-Darling test, should be used to decide if the sample could have been drawn from a normal distribution. The use of NPP and statistical tests can only provide a level of confidence that the population is normal—they do not guarantee that the population distribution is normal. However, it is likely that in some analyses the analyst will have to deal with non-normal distributions for small sample sizes. No matter the sample size (i.e., number of MC iterations), analysis results should provide the statistical accuracy of the reported results

whenever possible. Using the percentage error of the mean provides a normalized approach to both estimate the number of iterations and quantify the accuracy of the simulation results.

To summarize, this report has focused on 2 related topics: the number of MC iterations and the accuracy or error in the estimation of the mean of the probability distribution for the QI. For the majority of MC simulations, it is the estimation of this mean that is desired. These 2 topics are related through the CLT, and given one, the other can be determined when combined with the sample information for large sample sizes. Issues associated with small sample size have been mentioned and discussed in more detail for QI with a normal distribution.

7. References and Notes

1. "A technique in which a large quantity of randomly generated numbers are studied using a probabilistic model to find an approximate solution to a numerical problem that would be difficult to solve by other methods." Oxford Dictionaries [accessed 2015 Mar 20]. http://www.oxforddictionaries.com /us/definition/american_english/Monte-Carlo-method.

2. This step is somewhat vague because without knowing the value of π, it is not practical to determine whether or not the desired accuracy has been achieved. In practice, accuracy is generally quantified by the difference between successive approximations.

3. Robey RR, Barcikowski RS. The number of iterations in Monte Carlo studies of robustness. Paper presented at: the Annual Meeting of the American Educational Research Association; 1988 Apr 5–9; New Orleans, LA.

4. Hauck WW, Anderson S. A survey regarding the reporting of simulation studies. The American Statistician. 1984;38:214–216.

5. Devore JL. Probability and statistics for engineering and the sciences. 6th ed. Belmont (CA): Brooks/Cole; 2004. p. 238–241.

6. "Geometric probability involves the distributions of length, area, and volume for geometric objects under stated conditions. The same basic concept behind probability applies, but instead of calculating total outcomes and particular outcomes, calculate total area and particular area of a geometric figure using the formula *P = particular area/total area*." Chegg Inc. Definition of geometric probability [accessed 2015 Mar 18]. http://www.chegg.com /homework-help/definitions/geometric-probability-63.

7. Each input parameter may have a different pdf. The pdf is not necessarily a normal distribution, and, for example, could be a pdf consisting of a single value for an input parameter that has a constant value.

8. Devore JL. Probability and statistics for engineering and the sciences. 6th ed. Belmont (CA): Brooks/Cole; 2004. p. 286–294.

9. Devore JL. Probability and statistics for engineering and the sciences. 6th ed. Belmont (CA): Brooks/Cole; 2004. p. 294–297.

10. Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society. 1943;54:426–482.

11. Dunnigan K. Confidence interval calculation for binomial proportions. Fairfield (OH): Statking Consulting, Inc.; 2008. Report No.: P08-2008.

12. Agresti A, Brent C. Approximate is better than ‘exact’ for interval estimation of binomial proportions. The American Statistician. 1998 May;52(2):119–126.

13. Stein V. Confidence intervals for a binomial proportion. Statistics in Medicine. 1993;12:809–824.

14. Sauro J, James L. Estimating completion rates from small samples using binomial confidence intervals. Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting; 2005. p. 2100–2104.

15. Agresti A. An introduction to categorical data analysis. 2nd ed. (Wiley Series in Probability and Statistics). Hoboken (NJ): Wiley; 2007. p 9.

16. Devore JL. Probability and statistics for engineering and the sciences. 6th ed. Belmont (CA): Brooks/Cole; 2004. p. 294–296.

17. Devore JL. Probability and statistics for engineering and the sciences. 6th ed. Belmont (CA): Brooks/Cole; 2004. p. 291–292.

18. Devore JL. Probability and statistics for engineering and the sciences. 6th ed. Belmont (CA): Brooks/Cole; 2004. p. 299–301.

19. Wikipedia. Normality test. [accessed 2015 Apr 6]. http://en.wikipedia.org/wiki/Normality_test.

20. Razali N, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics. 2011;2(1):21–33.

21. Devore JL. Probability and statistics for engineering and the sciences. 6th ed. Belmont (CA): Brooks/Cole; 2004. p. 188–195.

22. Probability plots can be constructed assuming any population distribution. The idea is the same. If the plot of the sample data falls on a straight line, the data was drawn from a population with the specific distribution.

23. NIST/SEMATECH e-handbook of statistical methods. Washington (DC): National Institute of Standards and Technology; 2003 Jun 01. Normal probability plot. Section 1.3.3.21 [updated 2013 Oct 30; accessed 2015 Apr 7]. http://www.itl.nist.gov/div898/handbook/eda/section3/normprpl.htm.

24. MATLAB is a registered trademark of The Math Works, Inc., 3 Apple Hill Drive, Natick, MA, 01760.

25. Driels MR, Shin YS. Determining the number of iterations for Monte Carlo simulations of weapon effectiveness. Monterey (CA): Naval Postgraduate School (US); 2004. Naval Postgraduate School Report: NPS-MAE-04-005.

26. Driels MR, Shin YS. Determining the number of iterations for Monte Carlo simulations of weapon effectiveness. Monterey (CA): Naval Postgraduate School (US); 2004. Naval Postgraduate School Report: NPS-MAE-04-005. Eq. 13.

27. Driels MR, Shin YS. Determining the number of iterations for Monte Carlo simulations of weapon effectiveness. Monterey (CA): Naval Postgraduate School (US); 2004. Naval Postgraduate School Report: NPS-MAE-04-005. Eq. 14.

Appendix. MATLAB Code to Produce a Normal Probability Plot for Data in Array A

This appendix appears in its original form, without editorial change.

```
function makenormalplot = NorProbPlot(A)

mu = mean(A);
sd = std(A);
AS = sort(A);
N = length(A);
for i = 1:N
   x = (i-.5)/N;
   X(i) = norminv(x);
end

plot(X,AS,'.','MarkerSize',20)
title('Normal Probability Plot','FontSize',20)
xlabel('Z Percentile','FontSize',15)
ylabel('Observations','FontSize',15)
hold on
xx = [X(1) X(N)];
yy = [sd*X(1)+mu sd*X(N)+mu];
plot(xx,yy)
t1 = strcat('Mu =  ',num2str(mu));
text(-2,mu,1,t1)
t1 = strcat('StdDev =  ',num2str(sd));
text(-2,mu-1,1,t1)
hold off
```

1 (PDF) DEFENSE TECHNICAL INFORMATION CTR DTIC OCA

2 (PDF) DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT

1 (PDF) GOVT PRINTG OFC A MALHOTRA

9 (PDF) DIR USARL RDRL WML A W F OBERLE C PATTERSON M ARTHUR L STROHM T FARGUS R YAGER A THOMPSON B BREECH RDRL WML H J NEWILL

INTENTIONALLY LEFT BLANK.